CSC 2509 — Data Management Systems, Spring 2006

Assignment 1: SQL and Relational Algebra

Due Thu March 2
at the beginning of class

No late assignments will be accepted.

In problems 1–4, you are required to use the DB2 relational database management system. You are encouraged to use the DB2 system on CDF, but you may use any DB2 system available to you.[1] Be sure to store all your commands on a file, so that you can execute them for me if I ask you to. Each command should be tested using the DBMS, and its output should be recorded, either by running a Unix `script` session, or by dumping the output to a file. You should hand in the SQL commands and the output they produce. Each command and each output should be clearly marked with the question number that it corresponds to. Your solutions should be simple. Do not use subqueries unless you have to. Since most database systems come with good optimizers, you need not worry about efficiency. Your material should be well organized so that it is easy to grade, and the onus is on you to demonstrate that your SQL commands work correctly. 10% of the grade is for presentation. (You needn't use a document compiler like Latex or Troff, and no extra credit will be given for doing so.)

1. **Database Construction.** (16 points)

   (a) (5 points) Create a database with the tables listed below. Choose appropriate datatypes for each attribute.

   ```
   person[sin,name]
   address[sin,number,street,city,phone]
   doctor[sin,department,hospital,city,salary]
   patient[sin,sex,age,disease,bed,room,hospital,fee]
   treat[doctor_name,patient_name]
   ```

   Each relation and attribution has a straightforward meaning. For instance, in the following example

   ```
   person[12345,Amy]
   address[54321,21,Young,Toronto,4167771234]
   doctor[54321,tumor,sunnyBrook,Toronto,60000]
   patient[12345,female,54,brainTumor,307,1,sunnyBrook,43000]
   treat[Ken,Amy]
   ```

   person[12345,Amy] means Amy's sin number is 12345. The meaning of `address` is clear. `doctor[54321,tumor,sunnyBrook,Toronto,60000]` means doctor 54321 works in tumor department, sunnyBrook hospital in Toronto and his annual salary is $60,000. `patient[12345,female,54,brainTumor,307,1,sunnyBrook,43000]` means patient 12345 is a female of age 54, has disease brain tumor, is now in sunnyBrook hospital, room 307, bed 1 and her treatment fee is $43,000. `treat[Ken,Amy]` means doctor Ken treats patient Amy.

[1] However, we can only provide you with help and advice for the DB2 system on CDF. If you use another DB2 system and you run into trouble, you are on your own.

(b) (9 points) Populate the DB2 database of at least 30 tuples using the above five tables. Choose the data so that the queries in subsequent questions return non-trivial answers (i.e, not always an empty answer, and not always the same answer, not always a single answer). Moreover, be sure that the data satisfies realistic constraints as follows:

- no person has two addresses or two sin numbers,
- every person including doctors and patients mentioned in the database lives somewhere,
- there are at least 3 hospitals,
- each bed of any room can only be assigned to one patient,
- each room has no more than 3 patients,
- each patient only has one document file,
- each doctor only works in one hospital,
- for each doctor, he or she should be in the same hospital with all the patients treated by him or her,
- you may assume that a person's name uniquely identifies him or her.

You do not need to automatically verify these constraints or define any formal database constraints. Simply insert data into the database that satisfies the constraints.

(c) (2 points) Use SQL to show that for each doctor, he or she should be in the same hospital with all the patients treated by him or her.

2. **Basic SQL Queries** (7 points, 1 points each)
Express the following queries in SQL and pose them to the database constructed in Question 1. Of course, your queries should work for any reasonable set of data, not just the particular data in your database.

(a) Retrieve the sin and name of every patient.

(b) Retrieve the name and address of all people who live in `Toronto`.

(c) Retrieve the name and salary of all doctors who can treat disease `brainTumor` and who earn under $55,000.

(d) Retrieve the name, sex and age of every patient in room `307` at `sunnyBrook` hospital.

(e) Retrieve the name, age and phone number of every patient whose fee is more than $30,000.

(f) Retrieve the sin of every doctor who has at least one male patient older than 75.

(g) Retrieve the name of every doctor who lives and works in the same city.

3. **Complex SQL Queries** (12 points, 2 points each)
Express the following queries in SQL and pose them to the database constructed in Question 1. Of course, your queries should work for any reasonable set of data, not just the particular data in your database.

(a) Retrieve the name, room and hospital of every patient who is treated by more than one doctor.

(b) For every doctor, retrieve the doctor's sin and the sin of every patient treated by the doctor. Group the results by the doctor's sin.

(c) Retrieve the departments at `mountSinai` hospital in the city of `Toronto` that have the fewest number of doctors.

(d) For every hospital, retrieve the hospital name, doctor name, and department of every doctor who earns the *second* most in his department. Group the results by hospital name, and do not include doctors who work in the `heart` department.

(e) Retrieve the names of those patients who pay more than the average fee for all patients in the same hospital.

(f) For each doctor, retrieve the doctor's name and department and those rooms that have at least two of the doctor's patients.

4. **Database Update** (16 points, 2 points each)
Express the following updates in SQL and pose them to the database constructed in Question 1. If necessary, you may use a series of two or more SQL commands.

(a) Change the city of `Ken` from `Toronto` to `Oshawa`, and change his phone number to `(905)1111111`.

(b) Give a 1% raise to all doctors who work and live in different cities.

(c) Give 3% discount to all patients in a room that has more than 2 patients.

(d) Transfer patient `Linda` from her original hospital to another one, and change her doctor as well. Be sure all the constraints listed in Question1(b) are satisfied.

(e) Remove doctor `Ken` and his hospital information. (He is retiring.) Moreover, assign all his patients to some other existing doctors. Be sure all the constraints listed in Question1(b) are satisfied.

(f) Add a new patient to some hospital.

(g) Delete all the tuples in relation `person` who are neither doctors nor patients. (No trivial answers please)

(h) Delete all the tuples in relation `patient` who have disease `brainTumor`, and delete corresponding information from `treat`.

5. **Relational Algebra** (12 points, 2 points each)
Express every query in Question 2(b-g) in relational algebra. You may use an operator, $\rho_{A \leftarrow B}$, that changes the name of an attribute from B to A. That is, if T is a table, then $\rho_{A \leftarrow B} T$ is a table identical to T except that attribute B has been renamed as A.